

Collaborative Aspects of Collecting and Reflecting on Behavioral Data

Gabriela Marcu
University of Michigan
Ann Arbor, MI, USA
gmarcu@umich.edu

Allison N. Spiller
University of Michigan
Ann Arbor, MI, USA
aspill@umich.edu

ABSTRACT

Direct observation of behavior provides a unique type of data for reflecting on during a process of behavioral intervention. This study focuses on practitioners who specialize in operationalizing, recording, and monitoring behavior using data collection through paper-and-pencil or, increasingly, mobile computing. Applying an action research approach, we conducted fieldwork to understand observational data collection among practitioners providing children with special education support for behavioral needs. We present a model of collaborative data collection, which describes how practices are situated in the process of collecting data that are useful for reflection by teams of practitioners. We discuss how computer-assisted data collection could promote more systematic and rigorous practices, and design considerations for the collaborative aspects of collecting and reflecting on behavioral data. This study builds on research describing the practices of individuals who track their own behavioral data, and improves our understanding of informal documentation practices in organizations.

Author Keywords

Behavioral intervention; computer-assisted data collection; action research

CCS Concepts

•Human-centered computing → Empirical studies in collaborative and social computing; •Applied computing → Psychology;

INTRODUCTION

Understanding human behavior is a critical but challenging aspect of the social sciences. Researchers work to explain human behavior and advance knowledge. Individuals work to reflect on their own behavior and make healthy choices. A broad range of practitioners help individuals and society by combating adverse outcomes of behavior—such as physical ailments, mental health challenges, and encounters with the criminal justice system. At the core of these efforts is the difficult task of scrutinizing behavioral patterns and causes, in

order to effect change. Data are important for investigating and understanding behavior. But capturing accurate, reliable, and meaningful data on behavior is a nontrivial exercise.

Self-report of behavior provides a first-hand account and is particularly effective for cognitive processes that influence behavior. Limitations to self-report include the amount of self-awareness required, burden on the individual, and social acceptability bias. These issues are addressed in the design of tools to help individuals monitor and reflect on their own behavioral data, such as wearable devices that track physical activity [18], food diary apps that help monitor nutrition [28], and self-assessment for managing a mood disorder [31]. Personal informatics and consumer health informatics are areas focused on individual agency and autonomy in targeting one's own behavior using data-driven approaches.

Another approach involves an observer characterizing and documenting behavior. An observer can provide an external perspective on one's behavior that may be more objective, or provide different context and interpretation. For example, family members or cohabitants can identify behavioral patterns of which an individual may not be cognizant. Professionals can be trained to identify and characterize behaviors, then provide tailored feedback in line with goals. Direct observation of natural behavior is used in a broad range of contexts such as health care settings [36], sports [8], anthropology and ethology [21], and services marketing [17].

Compared to self-tracking, much less is known about how technology can support collecting data during observation of another's behavior. This paper revisits knowledge of personal informatics, to describe the collaborative aspects involved when observers can offer useful data on behavior. Li et al.'s model of personal informatics [25] outlines the cascading barriers that individuals face across five stages—we adapt these to the practices of observers. The *preparation* stage is deciding whether and how to collect data. Once the *collection* stage begins, there are challenges with using tools effectively and fitting in data collection with everyday practices. The *integration* stage involves preparing the data for reflection, by combining and organizing data, and creating useful visualizations. In the *reflection* stage, barriers are related to having holistic, consistent, and meaningful data in order to interpret behavior. Finally, the *action* stage involves challenging decision-making about how to act on the data—many systems do not make suggestions, but some support sharing data with others.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://www.acm.org/permissions).

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

© 2020 Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-6708-0/20/04 ...\$15.00.

DOI: <https://dx.doi.org/10.1145/3313831.3376623>

Chung et al. extend this model to account for the fact that much of this process is social and collaborative [11]. They suggest that technology could enable the collaborative generation of templates for collecting data in line with agreed upon goals, and provide tailored and goal-oriented visualizations for each stakeholder to review the data with the amount and type of detail they need. Marcu et al. similarly describe the need to support collaborative reflection among practitioners [32].

This paper bridges knowledge in personal informatics, and the evidence-based practice of behavioral intervention, to make three contributions:

- Understanding the work of practitioners specializing in behavioral intervention can inform the design of tools to improve systematic data collection by individuals irrespective of their knowledge or training.
- Applying personal informatics models to practitioners provides new insight into designing for informal documentation practices in organizational teamwork.
- A descriptive model, derived from how teams of practitioners collect and reflect on behavioral data, elucidates collaborative practices that can inform the design of tools which include multiple perspectives to understand behavior.

RELATED WORK

To this end, we focus on practitioners working in programs for children with behavior disorders. Data collection in this type of program supports "life-changing decisions ... includ[ing] residential placement, the use of restrictive behavioral procedures, changes or lack thereof in psychotropic medication, use of restrictive or labor intensive staffing, and so on" [51]. From an evidence-based perspective, data collection should not be disconnected from the intervention itself, yet practitioners have difficulty integrating these practices [44, 51]. We therefore draw from implementation science to investigate the challenges of implementing this type of evidence-based practice in everyday work [4].

Types of documentation in children's behavioral services

A range of documentation practices are used in the care of children across settings such as outpatient care, social services, and school [1, 32, 33, 43]. Saario et al.'s study of child health and welfare professionals [43] distinguished informal data collection practices from formal documentation used in communication across agencies. Formal documentation included the psychiatric daily record, medical case summary, or referral to another agency. Informal data collection included paper-based records that were used for "raw data in preparing" the formal electronic case record, including helping them understand what was going on with a child, or making a personal note of information that they were not ready to share because it could offend others.

Challenges with data collection in school-based services

Behavioral intervention is also an important part of school-based services, such as behavior management plans and special education supports. In these contexts, data collection is most effective in real time, to generate the progress monitoring

data that can be entered into the formal record (e.g., behavioral report cards, individualized education plans). Behavioral interventionists in early childhood programs, including teachers and therapists, "who collect and use child data demonstrate better instructional practices than those who do not" [44]. Increases in data collection have been shown to help interventionists refine behavioral goal setting they perform with each child [44]. In classroom settings, however, practitioners are working with multiple children at the same time, and this accounts for a number of barriers to implementing rigorous data collection in daily practices [34].

Studies also show that school-based practitioners value data collection, but they report rarely adhering to the practice and finding it difficult to do so [44]. When they do collect data, practitioners find themselves with a large number of data sheets that they do not analyze or use [34, 44]. Barriers to using collected data include a self-reported lack of knowledge or skills, a perception that the data are not meaningful, lack of time, and lack of tools [34, 44]. Practitioners express the need for support to integrate data collection practices into their work, as well as to analyze, interpret, and use the data [44]. When they have tried to adopt mobile applications to help with data collection and management, practitioners have had difficulty integrating them in their work [34].

APPROACH AND RESEARCH QUESTIONS

We conducted fieldwork using an action research approach, whereby practitioners examined their practices together with our research team [20]. That is, rather than maintaining the distance to study them as subjects of research, we used an action research process to combine their expertise with ours for deeper insights into their work. The action we took involved critiquing existing data collection practices, which were primarily paper-based, as well as prototyping and deploying a tablet application for computer-assisted data collection. Involving end users in reflecting on technology-in-use enabled us to explore how novel forms of computer-assisted data collection can be introduced democratically—for example, in ways that empower those accountable for the work of data collection [5]. This process was driven by two research questions:

RQ1: How do practitioners carry out activities of data collection alongside their other work?

RQ2: What are design considerations for integrating computer-assisted data collection in the work of practitioners?

We quickly learned that data collection in the moment occurs exclusively on paper. Yet we found computer-assisted data collection in the literature as far back as mobile devices afforded these practices. We consequently became interested in the lack of adoption in practice. To give context to findings from our action research process, we first recount the rich history of computer-assisted data collection.

DATA COLLECTION DURING REAL-TIME OBSERVATION

Researchers, especially those in the field of applied behavior analysis, have embraced the potential for computers to support real-time observation since the mid-1980s [22]. One of the earliest uses of mobile computing to aid in behavioral data

collection was by a research team studying the transfer of people with severe disabilities from institutions to group homes in Great Britain, who developed their own software to help measure the effects [41]. They used the Epson HX-20, considered the first truly mobile computer [48], and "its primary value in these studies was in collating and summarizing data quickly and without error" [41]. Before and after transfer to a group home, data were collected on clients' adaptive, maladaptive, and neutral behaviors in order to identify any changes. Data were also collected on staff member behaviors, such as instruction and guidance provided to clients.

Compared to traditional paper-and-pencil recording during observation, computer-assisted data collection has enabled the development of more complex analyses, for example, for "providing better understanding of the classroom social behavior of children and youth with EBD [emotional and behavior disorders]" [46]. Behavior in a classroom involves tightly interrelated factors such as student-teacher interactions, student-peer interactions within small or large groups, and a range of tasks and activities. Mobile computing gave way to the routine use of taxonomies such as Ecobehavioral Assessment Systems Software (EBASS), which combines data on 53 variables related to student behaviors, teacher behaviors, and ecology (current activity, task, or group structure) [16]. Based on their development of EBASS in the 1990s, Greenwood et al. reported that:

Initially, electronic technology replaced mechanical stop-watches and counters previously used to record the durations and frequencies of behavior. Currently, portable computers provide these and other functions, including observer training, calibration, data collection, case and database management, numerical analysis, and graphical displays.

In 2004, field researcher Gillian Ice assessed the state of mobile computing, comparing and contrasting it to the merits of paper and pencil [21]. Ice described paper and pencil as the preferred method when an existing taxonomy cannot be applied and possible states to be recorded are infinite. She also noted that computer adoption for data collection required extensive training. However, she and others encouraged the early adoption of computer-assisted data collection because it was likely to make processing the data less labor intensive, reduce missing data, and reduce human error resulting from complex protocols or data transfer from paper to computer [21, 50, 41]. Similarly, Harmann et al. noted "considerable merits of technology for behavior observation" while cautioning about downsides such as data overload, downtime during technology failure, and the need to standardize its use across studies to ensure generality [19].

The promise espoused by early adopters of these technologies has in some ways been realized. There is some evidence that computer-assisted data collection is more efficient than paper and pencil—due to time saved during data entry and analysis—with the same level of accuracy [49]. Computer-assisted data collection has thrived in applied behavioral research. Two DOS applications developed by researchers in the 1990s, The Observer [38] and the Multiple Option Observation System

for Experimental Studies (MOOSES) [50], have evolved into Windows and Android versions¹ that researchers use today. These systems are now considered foundational in observational research as an option for managing behavioral data [39], and have hundreds of citations (see [10, 27, 23, 37] for a range of recent studies that cite their use for data collection).

Data collection by practitioners

Literature on computer-assisted data collection by practitioners is much more limited. Early adopters Repp et al. began looking to computer-assisted data collection outside of research: "we see its value, however, not in providing grouped data for publication, but rather in a day-to-day program of evaluation through which staff are trained" [41]. Gravlee, Schulz, and colleagues used a community-based research approach, which employed local residents as neighborhood observers who completed structured checklists about the physical environment on handheld computers [15, 45]. Unexpectedly, they found that computer-assisted data collection had the benefit of enabling them to monitor training data. For example, they reviewed time stamps to understand the pacing of trainees, accessed the automatically calculated inter-observer agreement, and monitored certification requirements to determine when each observer completed their training [15]. The potential to design enhanced training processes has important implications for supporting the work of practitioners.

One example strictly from practitioners is that of industrial hygienists, whose work involves processing large amounts of information as they assess workplace environments for potential health hazards [2]. Malter and Davis were early adopters of handheld computers in their field [30]. As practitioners they noted that "we are being asked to do more with fewer resources. The handheld computer can help us do more with less" [30]. Their report detailing use of computer-assisted data collection, similar to literature already discussed, focused on the advantages of eliminating data transfer from paper records, and having access to data in real time:

Our primary goal was to eliminate the need to create paper records and the requirement to then transfer those written records into a desktop or notebook computer. Experience ... taught us we could also gain substantial benefits in data quality and 'real-time' learning by our field staff.

Lack of design exploration and guidance

The literature has focused on comparisons of computer-assisted data collection to traditional paper-and-pencil methods, with limited discussion around their design. In one exception, Couper has argued for usability evaluation to be applied as commonly as standard pilot testing of data collection materials such as paper surveys [12]. However, in the literature there is a lack of guidance on making design decisions before evaluating them. In health care settings, there is some evidence that the use of mobile devices for data collection—such as generating progress notes in a point-of-care patient record—can help practitioners save time, reduce errors, make

¹ See <https://www.noldus.com/the-observer-xt/gathering-data> and <http://mooses.vueinnovations.com/overview/mooses-overview>



Figure 1. Left: Practitioners collected data using a separate data sheet for each of the 12 children. Sometimes they could spread these out on a table to help record behaviors in real-time while observing. Right: Our Lilypad prototype mimicked existing practices by enabling navigation between all of the children’s data, and recording of behaviors using frequency counters and timers.

educated decisions, and provide more direct patient care [26]. But not much is known about how practitioners incorporate data collection into their everyday work, or how the design of computer-assisted data collection affects their work practices.

There has also been little focus on adapting to new interaction modalities such as mobile and touchscreen devices. Williams has investigated how easy-to-use interfaces, including touch-based interaction, can make assessment in the field accessible to those who are not experts or who have disabilities [53]. ECOVRD, or Event Coding and Visual Representation of Data, is one of the only computer-assisted data collection tools to leverage the affordances of a touchscreen, in this case on a tablet PC [13]. For example, each individual being observed is allocated a distinct space on the screen, and touching that part of the screen opens a pie menu with behaviors to record for that individual. This design approach is also unique compared to many other systems because it enables feasible navigation between multiple individuals under live observation.

SETTINGS AND METHODS

Our action research process unfolded over three phases. We performed the iterative steps of action research throughout this process: planning, taking action, and then reflecting on the results of the action [24]. During this long-term fieldwork, we took field notes and discussed them at weekly meetings in our lab. New insights were then verified back in the field with multiple practitioners who could provide differing perspectives. Findings were compared across participants and the literature. Inductive thematic analysis [7] was used by the researchers to reflect on outcomes from each phase of the fieldwork.

Phase I: Problem definition

Initial fieldwork took place in seven programs across four states in the U.S., which provide services for children and youth with disabilities. Four were school-based special education programs, two were residential programs, and one was a clinic providing therapeutic services. Across these programs, we performed 58 hours of observation and 62 semi-structured

interviews with practitioners such as educators, paraprofessional educators, social workers, behavioral therapists, and psychologists. Participants were recruited by word of mouth.

Together with these seven programs, we began to narrow our focus around the problem of incorporating data collection practices in their everyday work. Both the researchers and practitioners were interested in the lack of computer-assisted data collection methods, especially considering the availability and affordability of technology such as mobile computing. As we narrowed in on this focus, those practitioners still in the preparation stage who have not bought in to the usefulness of data collection, or have yet to begin these practices, naturally fell out of scope for our study. Our methods and findings are therefore aligned with the challenges experienced by those who have already started collecting data.

Phase II: Understanding practices

Once we defined the problem of data collection, we entered a phase of deeper fieldwork in partnership with one program. This new partnership was established based on a mutual desire to explore the introduction of computer-assisted data collection into everyday practices—which enabled us to focus on the stages of collection and reflection.

During this three month phase, we performed 61 hours of fieldwork. We visited the program twice a week and observed everyday activities surrounding collection and use of data (Figure 1, left). When possible, we conducted semi-structured interviews and contextual inquiry with the practitioners, to explore data collection within the context of their work [6].

When we came to understand the role that computer-assisted data collection could play in the work of practitioners, we began ideation and iterative prototyping in the lab [42]. Concepts and design decisions were verified with the practitioners during fieldwork visits. With continued iteration and refinement of features, we prepared what became the Lilypad tablet application (Figure 1, right), and decided together with the practitioners when it was ready for their use.

Phase III: Field deployment

We then conducted a field deployment, a method that provides "rich data about how closely a concept meets the target population's needs and how users accept, adopt, and appropriate a system in actual use over time" [47]. We deployed the Lilypad prototype over 11 months across two school years.

In the field deployment, we engaged a total of 20 practitioners working in special education and regular education, in four classrooms covering grades K-8 across three schools. Eleven of these practitioners were primary Lilypad users and their roles involved data collection as part of their everyday work. Nine had roles that did not involve collecting data, so they observed and reflected on use of Lilypad by their team members.

The deployment involved 554 hours of fieldwork including observation and informal interviews. As practitioners adopted Lilypad, we provided support through daily field visits for almost a week. Fieldwork then continued with visits once or twice per week for the duration of the deployment. Critical issues and bugs were reported by phone or email, and we provided support for technical issues such as those caused by updates to the operating system. Otherwise, our fieldwork engaged practitioners in reflecting on use of the prototype in the context of their individual and collaborative work. The prototype was not static during the deployment, and together we took advantage of opportunities that arose for adjusting the interface and functionality to better suit the needs of users.

Reflecting on the deployment, we began generating a first version of our model of collaborative data collection. We conducted interviews to validate and iterate on this model with the seven primary users of Lilypad, and the consulting behavior analyst who worked closely with our team throughout the process. Interviews were audio recorded and transcribed in full, then analyzed using deductive thematic analysis based on the elements in our model.

OVERVIEW OF BEHAVIORAL DATA IN PRACTICE

Comparing seven different programs in phase I, we noticed that their practices aligned with the five stages of personal informatics: preparation, collection, integration, reflection, and action [25]. We revisited this literature and confirmed that these stages were also descriptive of cases in which behavioral data is collected on others, rather than the self. Aside from self-awareness, the model of personal informatics focuses on many objective measures of behavior, for which sensing and other automation can be helpful. We thus found that the same process applies to an external observer who is noting behaviors objectively. In the context of observing others' behavior, we found the collection and reflection stages to have the most barriers, so we focus most on activities related to these stages.

In this section, we describe the type of behavioral data collected by practitioners, and for what purposes they wanted to use the data. We focus on the practices we found to be common across all of the programs we studied. Practitioners collected data on the frequency and duration of behaviors observed every day during school activities. Practitioners reflected on these data to monitor behaviors, and generate evidence of behavior for the purposes of reporting and formal

records. As previously mentioned, we found that practitioners generally used paper data sheets, and many of them were interested in computer-assisted data collection but had not found a tool that fit their practices.

A common challenge was the use of behavioral categories that were too broad to make the resulting data meaningful. These categories were typically generated as part of a classroom behavior management plan, to communicate to students clear behavioral expectations such as "use kind words" and "be safe". Although these categories can be effective in redirecting children's problem behaviors with verbal reminders of desired behaviors, their use as part of data collection practices did not produce specific data about, for example, how a child was displaying unsafe behavior (e.g., self harm, aggression toward peer, destruction of property). Therefore, *as practitioners integrated the use of data into the rest of their practices, there was a mismatch between the intervention of verbal feedback or redirecting behavior, and collecting data that would be useful for monitoring the progress of each child's behaviors.*

We gained an in-depth understanding of these practices during phases II and III, with one school district's set of practitioners who focus on identifying and supporting children with behavioral needs. Below we outline the collaborative aspects of how practitioners collected then reflected on data using both existing paper-based practices, and computer-assisted practices with our Lilypad prototype.

Collaborative aspects of the collection stage

We found that *practices of collecting data were tightly correlated with how the data would be used, and this was an inherently collaborative process* among teachers, paraprofessionals, social workers, and psychologists in the district. A teacher prepared all activities for their 12 students, and paraprofessionals worked alongside them in the classroom to provide behavioral support to students during activities. Children were identified for this type of support based on diagnoses of behavior disorders such as attention-deficit/hyperactivity disorder, obsessive compulsive disorder, and conduct disorder. A social worker directing the district's behavior disorder program oversaw these cases, creating individual behavior management plans for the year, managing student referrals, providing direct support especially with particularly challenging behaviors, and liaising with school administrators and parents. Finally, school psychologists and behavior analysts were contracted across the school district to provide consultation on cases as needed. Their role was to advise on implementing evidence-based strategies for managing behaviors.

The teacher and one or two paraprofessionals made up a classroom team. Together, they observed and managed children's behaviors throughout the school day. There was no dedicated team member who could focus their attention on data collection; instead, *all practitioners in the classroom incorporated data collection into their practices as part of their teamwork.* They prepared a data sheet for each student at the beginning of each school day, and then shared them based on which team member was working with each student during a particular activity. For example, during small group work each team

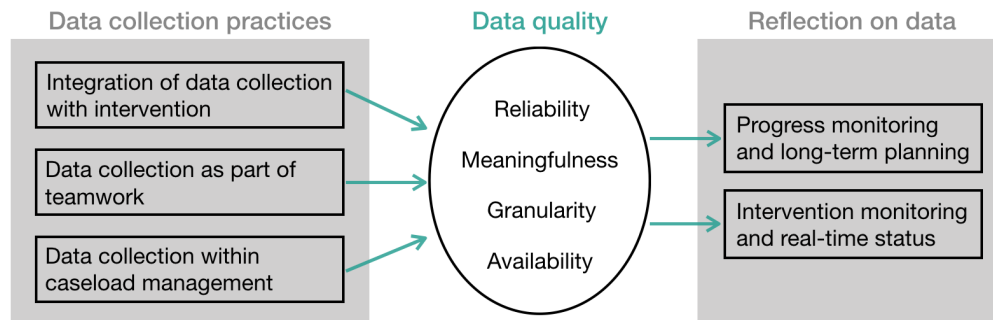


Figure 2. A model of collaborative data collection for behavioral intervention, describing data collection practices among practitioners (left), how these affected data quality (center), and how the resulting data were used for reflection (right).

member worked with a subset of the students at a table, keeping those students' data sheets on hand to record behaviors.

A child's data sheet contained a matrix, with eleven rows indicating each time period and activity in the day's schedule, and three columns of behavioral categories under which to record instances of behavior. In each cell of the matrix, practitioners entered a tally or start/end times of a behavior. The behavioral categories represented a classroom's behavior management plan, developed by the teacher with the social worker. Yet *each student also had an individual behavior management plan, and incorporating these into daily class-wide practices was nontrivial*. The classroom behavioral categories were used uniformly as a template for all of the student's data sheets, typically without tailoring them to individual behavior management plans. Therefore, practitioners sometimes noted additional details to characterize an individual student's behavior. However, due to the small amount of space available in the data sheet format, they used ad hoc shorthand to fit these notes in the two-inch cells, or along the margins.

Collaborative aspects of the reflection stage

Practitioners used data for reflection individually and collaboratively. *The data enabled them to monitor progress toward behavioral goals over time, and monitor how interventions are implemented on a daily basis*. These types of reflection were common across paper-based and computer-assisted practices. Immediate use was typically by sharing the data with a practitioner who was not present to observe certain behaviors. A member of the classroom team, who missed behaviors while they were momentarily out of the room, used the data as an update on the status of classroom work and dynamics. The social worker, who visited each classroom a few times a day, used the data to gain an understanding of a student's behavior throughout the day. The social worker could thus be informed during daily conversations they held with each student about individual behavioral goals and progress. By contrast, other practitioners such as psychologists and principals used the data on a long-term basis. Daily totals, graphs, and summarized reports were generated by classroom staff to help these other stakeholders reflect on the data for the purposes of monitoring progress over time and making ongoing treatment decisions.

With the design and deployment of the Lilypad prototype, our focus was on emulating existing practices, so that practitioners

could work with the same type of data they were used to, before we made any intentional changes. For example, Lilypad took the form of a tablet application because the practitioners were comfortable with the interaction paradigm and form factor, and its size was similar to the paper data sheets they used. Lilypad's interface centered around counters and timers for collecting frequency and duration data within the same time periods of the day's schedule. Still, *a transition to computer-assisted data collection naturally caused immediate changes in everyday practices*. Simple automation such as timestamps saved with each interaction altered the data available, though at first practitioners did not adjust to using this new information. Other functionality was used more immediately, such as the ability to check in on the status of a classroom or a particular student in real-time. The social worker made fewer spontaneous visits to each classroom, and instead checked in by reviewing the data on the tablet in his office.

A MODEL OF COLLABORATIVE DATA COLLECTION

Based on our understanding of how practitioners collect and use data for behavioral intervention, we developed a model of collaborative data collection. Illustrated in Figure 2, our model outlines the set of data collection practices we found among practitioners, how these affected the quality of the resulting data, and how these data were used for reflection. Li et al. [25] describe the cascading nature of barriers in this process, whereby barriers in early stages of data collection affect later stages such as reflecting and taking action on the data. Our model unpacks these barriers at the key stages of collection and reflection, highlighting the collaborative aspects of these activities among practitioners, and the challenges of achieving data quality in the context of behavioral intervention.

Data collection practices

We identified three practices of data collection that most affected the quality of the resulting data: integration of data collection with intervention; data collection as part of teamwork; and data collection within caseload management.

Integration of data collection with intervention

Our action research process served as a catalyst for practitioners to reflect on how their data collection practices were not always systematic and consistent, and explore how this might be addressed. Bruce, the program director, pointed to a lack of initial training on data collection (i.e., data taking)

as he reflected on two years during which they had tried both paper-based practices and our Lilypad prototype:

One of the challenges that we had is that there hadn't been any training on the data taking even with the paper and the pencil thing. So specifically in [one] room, one [practitioner] is really kind of hard-nosed about it, and [another has] a soft touch about it, so even within that grouping we had great variability [within the data].

Practitioners did not have operational definitions of behaviors accessibly documented, so that they could be consistent in how they collected data and responded to behaviors. As a result, practitioners in the same classroom did not reliably collect data on the same instances of behaviors, and did not give the same verbal feedback to students.

Practitioners were also not always in agreement about how to escalate their responses to behaviors. Behaviors were considered critical incidents if they required an escalated response such as removal from the classroom to help the student calm down (e.g., a time out), physically restraining a student who may harm themselves or others, or using the school's protocol for a disciplinary referral. A critical incident required additional paperwork to formally document the use of these interventions. However, practitioners did not have explicit criteria for what behaviors would be classified as critical incidents, and tended to make decisions on a case-by-case basis. As we implemented the Lilypad prototype into their practices, we engaged in long-term discussions around what criteria could be used to flag behaviors as critical incidents during data collection, as it was yet unclear how the technology should facilitate either intervention or documentation.

If the intervention did not follow evidence-based practice, the integration of data collection only enabled practitioners to continue straying from more effective strategies. Most of the classrooms we studied used punishment more than reinforcement strategies. Punishment is a focus on undesired behavior, and is reflected in data collection as taking away points as a consequence for behavior. In contrast, reinforcement is a focus on teaching and rewarding desired behavior, and is reflected in data collection as giving points for behavior. Research shows that punishment is less effective and can lead to adverse effects such as antisocial behavior—yet it is difficult for practitioners to avoid this strategy [29, 35]. The consulting behavior analyst, Christine, confirmed from her expert observation that this phenomenon was occurring during the course of our study. Inconsistent use of reinforcement strategies persisted during the Lilypad deployment, despite experimentation with functionality for both reinforcement or punishment procedures, through a counter button that could either count up or down. As with any type of behavior change, this finding highlights the difficulty of implementing consistent and effective behavioral reinforcement, even with computer-assisted practices.

Data collection as part of teamwork

The nature of teamwork within the classrooms required that data collection responsibilities be distributed across team members. Data collection and other practices were malleable due to the individualized and unpredictable nature of services for

children with behavioral challenges [32]. However, there was only one paper copy of a data sheet, which was difficult to share across the team. For example, one child's problem behavior led to his remaining in the classroom while the rest of the students went outside for recess. The team's malleability enabled any of the practitioners to stay with him, depending on the circumstances of that particular day. But the data collection tool was not as malleable as their practices, so they were not always able to distribute each child's data sheet to the appropriate practitioner accompanying them in that moment. When data sheets were difficult to pass around to the appropriate practitioner, this sometimes caused a lapse in data collection, and the practitioners relied on their memory to fill in the data later that day.

After the Lilypad deployment, practitioners discussed the change that occurred as a result of every member of the team being able to access the data collection tool at the same time on their own tablet. Kerry noted that it was easier for all team members to contribute to data collection, compared to their practices with paper data sheets: "only whoever had [each student's] point sheets would be ... taking that data, whereas now we can all take that data". Collaborative data collection therefore presented more opportunities to capture data on behaviors from multiple observers. We note, however, that this increased the challenge of maintaining consistent practices and reliable data, which we discuss in the data quality section.

Data collection within caseload management

In the classroom setting, observation is conducted on multiple individuals at the same time. Logistically, this was challenging for practitioners because they carried separate data sheets for each student. The nature of practitioners' collaboration therefore unfolded through a significant amount of what Bardram and Bossen describe as mobility work, when "mobility arises because of the need to get access to people, places, knowledge and/or resources" [3]. Practitioners would need to either lay the data sheets out on a flat surface (as seen in Figure 1), or search through all of the children's data sheets, which took their attention away from other work, as Jasmine recounted:

If you need to take points from on the go, it was kind of a pain in the neck, because we would be in the hallway and you have a stack of paper you are trying to flip through and take somebody's points while you are walking.

Conversely, with the use of Lilypad, Irene found it easier to navigate between the children's data using tabs in its interface. She was confident about being ready to record a behavior at any moment, including spontaneously when she encountered a student from another classroom in the hallway: "If I see an incident in the hallway, I'm like, 'I'm taking your points.' ... I'm like, 'Watch me. Here we go.'" Practitioners were able to use the interface to navigate quickly during a range of activities. Within the classroom, when students worked individually at their desks, practitioners could walk around to check students' work while carrying on data collection for all of them on their tablet.

Managing a number of cases at once required significant mobility to respond to the immediate needs of each child. For

example, transitions between activities tended to exacerbate challenging behaviors—whether changing from one activity to the next within the classroom space, or transferring the students to the school’s cafeteria, gymnasium, music room, or outdoors. During one of these transitions, challenging behaviors were more likely to emerge, and also more difficult to manage. Practitioners tended to bring data sheets on a clipboard, or Lilypad on a tablet, to be able to record any behaviors that occurred en route. However, with the demands of a transition, practitioners often forgot to bring these additional materials. As a result, behaviors during some of the most challenging parts of the day were not always captured accurately in the moment.

Data collection practices were not always well integrated into the classroom behavior management plan. As a core component of this plan, if students achieved a certain number of points, they received a reward at the end of the day. All practitioners and children in the classroom referenced this system throughout the day. As mentioned, the students were sometimes briefly in possession of their own data sheet as they were handing it off. Seeing their data could be upsetting to some students, as Kerry described:

[The students] would sometimes see on their [data sheet] that they lost a point early in the morning that they didn’t realize they lost, or they would see that someone else lost some points and point that out to them. So it definitely caused some problems.

Practitioners therefore tried to be discrete during data collection, but as Toby recounted, this was difficult when the data were on paper:

I would say, ‘give me the paper without looking at it’. Most of the time [they wouldn’t look at it], no. With Lilypad I didn’t have to worry about that because all I gotta do is hit this [home screen] button.

Similarly, Bruce felt that the use of Lilypad enabled more discrete data collection, which helped them with managing behaviors:

When we were doing paper and pencil, [the practitioner] would pick the pencil up to document the kids behavior and the kid would go off. Chairs would be flying around the room. So [Lilypad] is much more private, they don’t really know who is documenting what about what kid, so there’s been much less escalation of behaviors because of the way that the data is recorded.

Data quality

The combination of data collection practices described above affects four characteristics of the resulting data: reliability, meaningfulness, granularity, and availability.

Reliability

Practitioners find it difficult to maintain data collection practices throughout their day, especially when managing a caseload concurrently within the classroom setting. They are aware that their practices are not always consistent, but their focus is on capturing whatever data they can. As Vollmer

et al. [51] report, reliability measures are standard in applied research, but not among practitioners. Most commonly, inter-observer agreement is measured to find whether two observers agree that a behavior did or did not occur, thereby collecting the same data. Our findings indicate that without dedicated effort, time, or tools, it is difficult for practitioners to ensure reliability through such means.

As previously noted, the use of computer-assisted data collection to distribute responsibility among the team can introduce new risks to reliability. However, our deployment also revealed that these tools could be leveraged to promote reliability through team awareness. We found that during use of the Lilypad prototype, practitioners were so aware of concurrent data collection by multiple team members, that they went out of their way to prevent against duplicate data. They appropriated Lilypad’s interface to mediate the risk by using the back button to refresh the data, as described by Kerry:

A lot of times I will back it out and go back in so that I know that it is completely refreshed and ... I see that no one took a body [control] point from student C in the last 20 minutes, and he has been [wandering around] all over the place, then I will go ahead and take that [point].

Given the existing teamwork among practitioners, simple visual cues within an interface could be used to increase awareness of when and how others have performed data collection.

Meaningfulness

Our focus on teamwork revealed that the motivation to collect data is often based on how meaningful the data will be to others, rather than for oneself. This finding deepens our understanding, since past work such as Sandall et al.’s [44] has reported that practitioners do not find data meaningful because they do not feel the data give them a better understanding than what they can observe and intuit themselves in their daily work with children. Our study also shows that even when practitioners are interested in data collection because they are bought into the idea of the data being useful, as many of our participants were, their efforts did not always result in data that were meaningful.

We highlight how the affordances of data collection tools affect the meaningfulness of the resulting data, regardless of the practitioners’ motivation. For example, the Lilypad prototype included expandable textboxes that did not limit the amount of space available for notes. Kerry reflected on how much more challenging it had been to fit notes in the margins of a piece of paper: “it was a mess, you would run out of room on the sides of the paper, you could only write so much”. The ability to recreate events later on was important for the team to be able to reflect and make informed decisions. They were not looking for more quantitative evidence of a behavior, instead they were looking for the qualitative notes that could give an account of the behavior and its contextual details. With more room to write notes, Bruce explained how the practitioners felt they captured data that told a more complete story:

“It was nice to show administration what was really going on. The [notes] piece is critical. When they took a look at all of what he said, things started to make sense. I

showed them on the iPad and was able to scroll the full day and show the [notes]."

Granularity

The granularity of the quantitative data—or the level of detail with which data were captured—is also limited by the tools used. Practitioners prepare the data for reflection through manual transfer onto graphs or reports, similar to Li et al.'s description of the integration stage for individuals [25]. However, this process is complicated by multiple practitioners managing multiple cases. We found that with a dozen children per classroom, this would amount to 84 data sheets per week, handled by three practitioners. Consequently, daily totals were used most, at a loss of granularity such as having data broken down by period, as well as the accompanying qualitative notes.

Computer-assisted data collection with Lilypad immediately enabled practitioners to increase the granularity with which they collected and used data. However, adjusting to this ability took time because it required a change in their mental model of the data. They were accustomed to thinking of points in terms of daily totals, for example, so when we were designing Lilypad together, they felt that we were over-complicating functionality related to how data would be demarcated by period. For example, they did not feel strongly about whether Lilypad should automatically advance to the next period based on the time, or allow the user to manually advance to the next period according to the actual activity, to ensure the points would be associated with the correct activity regardless of whether the class was running on schedule.

After the experience of using Lilypad over time, they appreciated the granularity with which they could reflect on the data. Anthony explained how he identified patterns and even discussed them with students:

We're actually able to see at what intervals or what times of the day [students] have issues, like - 'hey, you do well during Language Arts, but for some reason your points are always dropping when math comes'. It's stir[ed] up a conversation, like 'what's going on?' [And the student said] 'I really don't get it' or 'I really don't like math'.

Availability

Data must be available to others in order for them to access it and reflect on it. On paper, practitioners need to make a copy of the data available to others. As has been reported, a key motivation for employing computer-assisted data collection is to streamline the workflow from collecting to using the data. Within Li et al.'s model [25], this means shortening the integration stage between collection and reflection, by facilitating or automating activities such as combining, organizing, and graphing data. With the Lilypad prototype, the integration stage was shortened and Irene explained the utility of data being accessible to others in real time: "When I ... take the point, it's all right there and it's all on her computer. So, it's made my workload a lot less".

After collecting data with Lilypad on their tablets, the practitioners would use laptop and desktop computers to incorporate the data into the formal record, or reports that were emailed to stakeholders such as parents or administrators. Shortening

the integration stage not only improves availability for those who need access to the data, but also changes the accountability of the practitioner performing data collection—often not someone with a lot of power within the organization. As we discussed additional functionality supporting the transfer of data from Lilypad onto reports, Anthony discussed this accountability: "That would be cool because that's a way of covering ourselves ... I gotta get this done since I'm the one who input this data, I gotta have these filled out".

Reflection on data

The above quality characteristics of the data collected then influence how the practitioners can perform reflection on data. We describe practices around two purposes for reflecting on data: to monitor progress toward behavioral goals and adjust long-term plans accordingly; and to monitor how interventions are implemented on a daily basis across practitioners, including providing a status in real-time across a team.

Progress monitoring and long-term planning

The ability to monitor progress toward behavioral goals was a key motivator in collecting data. As already mentioned, our findings align with Sandall et al.'s [44] in that practitioners felt they did not need data collection to have a good understanding of each child's progress. Instead, we found that practitioners were motivated to show evidence of a child's behavior to those who were not in the classroom every day. For example, an administrator was hesitant to change a child's placement, so the classroom team collected more thorough data to show that the child's behavioral needs could no longer be met by their program. Parents sometimes had a difficult time understanding that their child's behavior could be different at school compared to what they saw at home. The classroom team documented one child's inappropriate language by transcribing it into Lilypad in real time, to have a record for the parent.

As practitioners gained experience with using Lilypad, however, they found that it gave them new capabilities for identifying behavioral patterns they had not otherwise noticed. For example, Bruce explained how they discovered that behaviors were different in the classroom than when the students moved to a period that took place in a different room (e.g., music room, gymnasium), which they referred to as specials:

When the students were in specials with just the two [paraprofessionals] and the classroom teacher wasn't physically present, we [saw a] spike [in] ... a lot of behaviors. So that was important information for us to discern or separate, how much of the behaviors are happening in specials and how much are in the regular classroom.

Intervention monitoring and real-time status

Data were also used to monitor how behavioral intervention was unfolding on a given day, and provide a metric for gauging a child's state in a given moment. For example, Bruce's role as the program director included supporting the intervention each classroom team provided by removing a child from a classroom to deescalate behavior, and give on-on-one coaching in his office. During the Lilypad deployment, he had his own tablet which he used to periodically review data and check if any of the children had escalating behaviors. He felt

this improved his ability to intervene at appropriate moments, compared to the paper data sheets which did not give him real-time access to data:

I would get [the data sheets] at the end of the day, and I would kind of review them ... It would help me know what to look for the next day, but it's too little too late in many circumstances.

Students were also interested in knowing their status in real time. They tended to look over to a practitioner's clipboard or tablet, or ask for updates on their points for the day. Managing a dozen students' cases made it difficult for practitioners to provide frequent feedback to all of them on their data, which is key to effective behavioral intervention [52]. The purpose of the daily point system was to motivate students and give them frequent feedback on their behavior. Therefore, their interest in seeing their data was a positive aspect of this classroom environment. This finding suggests that data collection tools could be designed to facilitate reflection by the students themselves, in addition to the various practitioners.

DISCUSSION

Our focus on the work of practitioners specializing in behavioral intervention reveals that they experience the same pitfalls common among individuals who perform personal data collection for self-monitoring: (1) they collect data on too many items, (2) they do not collect enough data on triggers and context, and (3) they conduct experimentation without scientific rigor [9]. Practitioners suggested a need for training on data collection practices, however past work has indicated that training is not necessarily effective [44]. Computer-assisted data collection, however, presents an opportunity to scaffold training and provide ongoing supports for rigorous data collection. *Interactive operational definitions of distinct behaviors or behavioral categories can help users clearly define, refine, and adhere to what exactly they are observing.* Users should be able to reference operational definitions on demand, for example through a hyperlinked behavior with a popup or tooltip providing its definition. Users should also be able to clarify definitions with others or propose edits, for example using wiki-style collaborative editing.

Behavioral intervention is an evidence-based practice, but classroom settings are exemplary of messy real-world contexts where rigorous implementation is difficult. For example, we explored the tendency to focus on problem behavior (with responses that can be characterized as punishment) rather than the more effective approach of reinforcing desired behavior. Personal informatics has recently focused on promoting rigorous experimentation. Choe et al. [9] argue that personal data collection is driven by self-experimentation, and tools should be designed to introduce more scientific rigor into their process—for example, by providing a platform that follows conventions of research design such as baseline assessment and variability in data. We found that simply providing flexibility within computer-assisted data collection, in the form of a counter button that can give points or take away points, did not even result in practitioners experimenting with new practices. This suggests that *scaffolding evidence-based practices and providing some restrictions and forcing functions may be*

more effective than enabling flexibility in use. This approach could also be used to promote reliability, for instance through inter-rater reliability.

Our model of collaborative data collection can be used to guide design and implementation of computer-assisted data collection, so that resulting data have qualities that are useful for reflection—reliability, meaningfulness, granularity, and availability. This contribution complements the literature describing how behavioral data can be represented and shared. For example, Chung et al. suggest that technology could enable the collaborative generation of templates for collecting data in line with agreed upon goals, and provide tailored and goal-oriented visualizations for each stakeholder to review the data with the amount and type of detail they need [11]. In the design of tools for self-monitoring, Choe et al. recommend supporting reflection through the inclusion of unstructured, open-ended data and the ability to share data with others [9]. We add a perspective from long-term action research including investigation of technology in use.

Our study also has implications for informal documentation among a range of health care and social service practitioners. A common form of data collection involves updating and managing the client record. Electronic records are now commonplace, but a range of practices are used more informally to collect and manage data in ways that help practitioners reflect on the client's immediate status and long-term trajectory—often through a combination of computers and paper [14, 40]. *Our model can inform design and implementation in organizational settings, because it elucidates how practitioners incorporate informal documentation practices into their everyday work with multiple team members, while managing multiple concurrent client cases.* We found that data collection practices for behavioral intervention have significant similarities with the universal aspects of informal documentation Park et al. found enacted across different health care settings [40].

CONCLUSION

We have presented a model of collaborative data collection, drawn from the experiences of practitioners who specialize in behavioral intervention in their everyday work. We used an action research approach to engage practitioners in over 100 hours of fieldwork across eight different programs serving children with behavioral needs. Together, we investigated paper-based and computer-assisted data collection, including our own Lilypad prototype. This study improves our understanding of informal documentation practices by drawing from extensive work studying the practices of individuals who track their own health data. Our design insights also have implications for computer-assisted data collection in support of treatment and quality of life for individuals with disabilities.

ACKNOWLEDGMENTS

We are grateful to our partners and participants in this work. We thank UMSI colleagues and our reviewers who provided helpful feedback on drafts of this paper. This material is based upon work supported by the National Science Foundation under Grant No. IIS 1816319.

REFERENCES

- [1] Ofra Amir, Barbara J. Grosz, Krzysztof Z. Gajos, Sonja M. Swenson, and Lee M. Sanders. 2015. From care plans to care coordination: Opportunities for computer support of teamwork in complex healthcare. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1419–1428.
- [2] American Industrial Hygiene Association. 2019. Discover Industrial Hygiene. (2019). Retrieved August 7, 2019 from <https://www.aiha.org/about-ih/Pages/default.aspx>
- [3] Jakob E Bardram and Claus Bossen. 2005. Mobility work: The spatial dimension of collaboration at a hospital. *Computer Supported Cooperative Work (CSCW)* 14, 2 (2005), 131–160.
- [4] Mark S. Bauer, Laura Damschroder, Hildi Hagedorn, Jeffrey Smith, and Amy M. Kilbourne. 2015. An introduction to implementation science for the non-specialist. *BMC Psychology* 3, 1 (2015), 32.
- [5] Marc Berg. 1998. The politics of technology: On bringing social theory into technological design. *Science, Technology, & Human Values* 23, 4 (1998), 456–490.
- [6] Hugh Beyer and Karen Holtzblatt. 1997. *Contextual Design: Defining Customer-Centered Systems*. Elsevier.
- [7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.
- [8] Julen Castellano, Abigail Perea, Lorea Alday, and Antonio Hernández Mendo. 2008. The measuring and observation tool in sports. *Behavior Research Methods* 40, 3 (2008), 898–905.
- [9] Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. Understanding quantified-selfers’ practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1143–1152.
- [10] Jason C. Chow and Joseph H. Wehby. 2019. Profiles of problem behavior in children with varying language ability. *Journal of Emotional and Behavioral Disorders* 27, 2 (2019), 110–118.
- [11] Chia-Fang Chung, Kristin Dew, Allison Cole, Jasmine Zia, James Fogarty, Julie A. Kientz, and Sean A. Munson. 2016. Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient-generated data. In *Proceedings of the Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 770–786.
- [12] Mick P. Couper. 2000. Usability evaluation of computer-assisted survey instruments. *Social Science Computer Review* 18, 4 (2000), 384–396.
- [13] Ute Fischer, Deepak Jagdish, and Abbas Attarwala. 2010. ‘ECOVRD’—A Tablet PC-Based Tool to Support Observational Studies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 54. SAGE Publications, Los Angeles, CA, 631–635.
- [14] Geraldine Fitzpatrick. 2004. Integrated care and the working record. *Health Informatics Journal* 10, 4 (2004), 291–302.
- [15] Clarence C. Gravlee, Shannon N. Zenk, Sachiko Woods, Zachary Rowe, and Amy J. Schulz. 2006. Handheld computers for direct observation of the social and physical environment. *Field Methods* 18, 4 (2006), 382–397.
- [16] Charles R. Greenwood, Judith J. Carta, Debra Kamps, Barbara Terry, and Joseph Delquadri. 1994. Development and validation of standard classroom observation systems for school practitioners: Ecobehavioral Assessment Systems Software (EBASS). *The Council for Exceptional Children* 61, 2 (1994), 197–210.
- [17] Stephen J. Grove and Raymond P. Fisk. 1992. Observational data collection methods for services marketing: an overview. *Journal of the Academy of Marketing Science* 20, 3 (1992), 217–224.
- [18] Daniel Harrison, Paul Marshall, Nadia Bianchi-Berthouze, and Jon Bird. 2015. Activity tracking: barriers, workarounds and customisation. In *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 617–621.
- [19] Donald P. Hartmann, Billy A. Barrios, and David D. Wood. 2004. Principles of Behavioral Observation. In *Comprehensive Handbook of Psychological Assessment, Vol. 3. Behavioral Assessment*. John Wiley Sons Inc., Hoboken, NJ, USA, 108–127.
- [20] Gillian R. Hayes. 2011. The relationship of action research to human-computer interaction. *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, 3 (2011), 15.
- [21] Gillian H. Ice. 2004. Technological advances in observational data collection: The advantages and limitations of computer-assisted data collection. *Field Methods* 16, 3 (2004), 352–375.
- [22] SungWoo Kahng and Brian A. Iwata. 1998. Computerized systems for collecting real-time observational data. *Journal of Applied Behavior Analysis* 31, 2 (1998), 253–261.
- [23] Elin Koppelaar, Hanneke J.J. Knibbe, Harald S. Miedema, and Alex Burdorf. 2012. The influence of ergonomic devices on mechanical load during patient handling activities in nursing homes. *Annals of Occupational Hygiene* 56, 6 (2012), 708–718.
- [24] Kurt Lewin. 1946. Action research and minority problems. *Journal of Social Issues* 2, 4 (1946), 34–46.
- [25] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 557–566.

- [26] Yen-Chiao Lu, Yan Xiao, Andrew Sears, and Julie A. Jacko. 2005. A review and a framework of handheld computer adoption in healthcare. *International Journal of Medical Informatics* 74, 5 (2005), 409–422.
- [27] Erika Lunkenheimer, Nilam Ram, Elizabeth A. Skowron, and Peifeng Yin. 2017. Harsh parenting, child behavior problems, and the dynamic coupling of parents' and children's positive behaviors. *Journal of Family Psychology* 31, 6 (2017), 689.
- [28] Yuhan Luo, Peiyi Liu, and Eun Kyoung Choe. 2019. Co-Designing food trackers with dietitians: Identifying design opportunities for food tracker customization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 592.
- [29] John W. Maag. 2001. Rewarded by punishment: Reflections on the disuse of positive reinforcement in schools. *Exceptional Children* 67, 2 (2001), 173–186.
- [30] David E. Malter and Terry J. Davis. 2003. Data collection and "real-time" learning using handheld computers. *Applied Occupational and Environmental Hygiene* 18, 5 (2003), 321–330.
- [31] Gabriela Marcu, Jakob E. Bardram, and Silvia Gabrielli. 2011. A framework for overcoming challenges in designing persuasive monitoring and feedback systems for mental illness. In *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare*. IEEE, 1–8.
- [32] Gabriela Marcu, Anind K. Dey, Sara Kiesler, and Madhu Reddy. 2016. Time to reflect: Supporting health services over time by focusing on collaborative reflection. In *Proceedings of the Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 954–964.
- [33] Gabriela Marcu, Allison Spiller, Jonathan Arevalo Garay, James E. Connell, and Laura R. Pina. 2019. Breakdowns in home-school collaboration for behavioral intervention. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 661.
- [34] Gabriela Marcu, Kevin Tassini, Quintin Carlson, Jillian Goodwyn, Gabrielle Rivkin, Kevin J. Schaefer, Anind K. Dey, and Sara Kiesler. 2013. Why do they still use paper?: Understanding data collection and use in Autism education. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3177–3186.
- [35] G. Roy Mayer. 1995. Preventing antisocial behavior in the schools. *Journal of Applied Behavior Analysis* 28, 4 (1995), 467–478.
- [36] Nicholas Mays and Catherine Pope. 1995. Qualitative research: Observational methods in health care settings. *BMJ* 311, 6998 (1995), 182–184.
- [37] Anna Schmidt Naylor, Debra Kamps, and Howard Wills. 2018. The effects of the CW-FIT group contingency on class-wide and individual behavior in an urban first grade classroom. *Education and Treatment of Children* 41, 1 (2018), 1–30.
- [38] L.P.J.J. Noldus. 1991. The Observer: a software system for collection and analysis of observational data. *Behavior Research Methods, Instruments, & Computers* 23, 3 (1991), 415–429.
- [39] Jamie M. Ostrov and Emily J. Hart. 2014. Observational methods. In *The Oxford Handbook of Quantitative Methods, Volume 1: Foundations*, Todd D. Little (Ed.). Oxford University Press, Inc., New York, NY, USA, 286–304.
- [40] Sun Young Park, Katie Pine, and Yunan Chen. 2013. Local-universality: Designing EMR to support localized informal documentation practices. In *Proceedings of the Conference on Computer Supported Cooperative Work*. ACM, 55–66.
- [41] Alan C. Repp, Kathryn G. Karsh, Rick Van Acker, David Felce, and Martin Harman. 1989. A computer-based system for collecting and analyzing observational data. *Journal of Special Education Technology* 9, 4 (1989), 207–217.
- [42] Yvonne Rogers, Hellen Sharp, and Jeni Preece. 2011. *Interaction Design: Beyond Human Computer Interaction* (3rd ed.). John Wiley & Sons Ltd., Chichester, West Sussex, UK.
- [43] Sirpa Saario, Christopher Hall, and Sue Peckover. 2012. Inter-professional electronic documents and child health: A study of persisting non-electronic communication in the use of electronic documents. *Social Science & Medicine* 75, 12 (2012), 2207–2214.
- [44] Susan R. Sandall, Ilene S. Schwartz, and Betsy Lacroix. 2004. Interventionists' perspectives about data collection in integrated early childhood classrooms. *Journal of Early Intervention* 26, 3 (2004), 161–174.
- [45] Amy J. Schulz, Shannon N. Zenk, Srimathi Kannan, Barbara A. Israel, Mary A. Koch, and Carmen A. Stokes. 2005. CBPR approaches to survey design and implementation. *Methods in Community-Based Participatory Research for Health* (2005), 107–27.
- [46] Richard E. Shores and Joseph H. Wehby. 1999. Analyzing the classroom social behavior of students with EBD. *Journal of Emotional and Behavioral Disorders* 7, 4 (1999), 194–199.
- [47] Katie A. Siek, Gillian R. Hayes, Mark W. Newman, and John C. Tang. 2014. Field deployments: Knowing from using in context. In *Ways of Knowing in HCI*. Springer, 119–142.
- [48] Tony Smith. 2007. From 1981: the World's first UMPC; Epson's battery powered mobile computer. *The Register* (3 May 2007). https://www.theregister.co.uk/2007/05/03/forgotten_tech_epson_hx20/, last accessed on 08/01/19.

- [49] Jon Tapp, Renata Ticha, Erin Kryzer, Meaghan Gustafson, Megan R. Gunnar, and Frank J. Symons. 2006. Comparing observational software with paper and pencil for time-sampled data: A field test of Interval Manager (INTMAN). *Behavior Research Methods* 38, 1 (2006), 165–169.
- [50] Jon Tapp, Joseph Wehby, and David Ellis. 1995. A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, & Computers* 27, 1 (1995), 25–31.
- [51] Timothy R. Vollmer, Kimberly N. Sloman, and Claire St. Peter Pipkin. 2008. Practical implications of data reliability and treatment integrity monitoring. *Behavior Analysis in Practice* 1, 2 (2008), 4–11.
- [52] Richard M. Wielkiewicz. 1995. *Behavior Management in the Schools: Principles and Procedures*. ERIC.
- [53] Drew Marie Williams. 2018. *Developing Accessible Collection and Presentation Methods for Observational Data*. Ph.D. Dissertation. Marquette University, Milwaukee, WI, USA. https://epublications.marquette.edu/dissertations_mu/836